

Leader

Benchmarking and the percentile assessment of RA: adding a new dimension to rheumatic disease measurement

It is somewhat surprising that the importance of benchmarking¹ was not generally realised until the last decade.²⁻⁵ Benchmarking conveys extraordinary advantages in the understanding of clinical trials, observational studies, and ordinary clinic patients by providing guidelines of expected values. Benchmarks allow us to assess the *status* of patients, whereas most clinical trials are concerned only with *change*. *Status* tells us how patients are doing today and predicts short and long term outcomes, including costs, work disability, and death. *Change* describes treatment efficacy, but has little to do with prediction or outcome. In the long run, it is *status* not *change* that is important.

Benchmarking requires a relevant, generalisable, large, well characterised population sample that is representative of either a population based or a clinical based cohort of patients with rheumatoid arthritis (RA). In the former instance, milder cases and cases not currently meeting RA may be included. The clinical sample should be composed of cases typical of those found in clinical practice, and should be representative of clinical practice in general. In addition, the sample must have enough detail to ensure that relevant RA covariates are captured. Duration of disease can be important for outcomes such as work disability, and mortality can be directly pegged to disease duration. Finally, to be useful the sample must be large enough so that there can be confidence that the estimates are reliable.

Benchmarking also requires accurate and reliable measures, of which there are now a wide range that are relevant to rheumatology and can form the basis of the benchmarks. In addition to the Health Assessment Questionnaire (HAQ),⁶ pain scale, and Short Form-36 (SF-36) subscales^{7,8} of the current report, other relevant scales include patient global severity, fatigue, swollen and tender joint counts, and laboratory tests such as the C reactive protein or erythrocyte sedimentation rate.

The study of Wiles *et al*⁹ is one more in a series of superior reports from the Norfolk Arthritis Register community based inception cohort study designed by Deborah Symmons and her colleagues in Manchester.¹⁰⁻¹² It meets the two criteria suggested above, and delineates population based RA/polyarthritis from clinical RA as represented by hospital patients and/or those meeting ACR RA criteria at onset. One limitation of this study is that the sample size is small, a limitation that is related to the low incidence of RA in the community; and this limitation may be reflected in a lack of precision of the median estimate. Although the authors do not provide confidence intervals for their data, they can be estimated from similar data from community rheumatologists available in the US National Data Bank for Rheumatic Diseases (NDB).⁵ For 213 patients with RA with a duration of RA of 4.75-5.25 years, the median and bootstrapped 95% CI for HAQ is 1.0 (95% CI 0.875 to 1.25), or an upper and lower CI 0.125 units above or below the median. For the SF-36 physical function scale those values are 55 (95% CI 40 to 60).

With such excellent data in hand, the authors propose that their five year results should be used as a measure

against which five year outcomes from other cohorts may be judged. However, this raises the important question as to whether benchmarks at specific time intervals are required for patient self reported data. We have shown elsewhere that HAQ scores¹³ and other status measures¹⁴ change little over time. To consider the issue further in the variables under study by Wiles *et al*,⁹ we randomly selected 12 949 patients with RA from the NDB in the years 1999-2001. The NDB is a databank of American patients with arthritis obtained from the practices of 244 (for the current report) rheumatologists.⁵ Using quantile (median) regression, we note that HAQ scores increase 0.0133 (SE.001) units a year. Put into perspective, the median values (and their implied confidence intervals) presented by Wiles *et al* might be expected to encompass HAQ values ± 9.4 years; or through a disease duration of 0 to 13.4 years. A similar case for limited change with time can be made for SF-36, where, for example, the physical function score (NDB n=8479) decreases at -0.578 (SE 0.072) units a year. For pain (0-10 scale, NDB n=12 870), the change is 0.021 (SE 0.003) units a year, a change that is essentially clinically meaningless. In addition, there is very little time related change in the other SF-36 variables and in the other clinical measures shown in table 1. This relative stability over time occurs because it is inflammation and/or pain rather than disease duration that is the primary driving force for most of these clinical measures. These observations suggest that it is not necessary to benchmark at specific time points in the duration of RA for the measures included in table 1.

Table 1 Health status scores of patients with RA in the practices of 244 American rheumatologists*, as registered in the US National Data Bank for Rheumatic Diseases (1999-2001)

Variable	n	25th Percentile	Median	75th Percentile
HAQ†				
Enrollees	8 399	0.375	1.00	1.625
All patients	12 949	0.375	1.00	1.625
Pain (0-10)				
Enrollees	8 387	1.5	3.0	6.0
All patients	12 870	2.0	4.0	6.5
Patient global (0-10)				
Enrollees	8 355	1.0	2.50	5.0
All patients	12 761	1.5	4.0	5.5
Fatigue (0-10)				
Enrollees	8 330	1.5	4.0	7.0
All patients	12 164	2.0	4.5	7.0
SF-36				
PCS‡	8 479	26.6	32.8	40.2
MCS‡	8 479	34.0	46.5	55.5
Physical function	8 479	25.0	50.0	75.0
Role-physical	8 311	0.0	25.0	100.0
Bodily pain	8 351	31.0	41.0	42.0
General health	8 331	50.0	57.0	65.0
Vitality	8 352	30.0	45.0	60.0
Social functioning	8 362	50.0	75.0	100.0
Role emotional	8 282	33.3	100.0	100.0
Mental health	8 346	64.0	80.0	88.0

*Data use one randomly selected observation for each patient. Enrollees are patients agreeing to participate in long term RA surveys. "All patients" are participants plus non-participants. Higher scores indicate better status for SF-36 measure. For other measures lower scores indicate better status.

†HAQ = Health Assessment Questionnaire; PCS = physical component score; MCS = mental component score.

Although time-specific benchmarks of patient self reported data may be more accurate than those that ignore the time element, they have the important limitation that they can only be used at a single time point, thereby substantially limiting their clinical usefulness. However, it is clearly necessary to account for time when outcomes such as work disability, total joint arthroplasty, and mortality are concerned.¹⁵⁻¹⁸

How should benchmarks be used? The data of table 1 represent a large and generalisable sample of patients with RA, making it possible to apply this table to RA studies and patients. In addition, more detailed charts with values at 10 percentile units or less are available at www.arthritis-research.org. If these data are applied to the results of randomised controlled trials (RCTs) of the anti-tumour necrosis factor agents infliximab¹⁹ and etanercept²⁰, we note that before starting treatment HAQ scores were 1.7 and 1.65, corresponding to percentile values of 77 and 76, respectively. At study closure, these percentiles had changed to 59 and 55, respectively. These percentile data give additional insight into the effectiveness of these agents that is not apparent from ACR improvement scores or raw HAQ scores, for they put into perspective the results of patients treated with these agents in comparison with all other patients with RA. In addition, percentile benchmarks allow one to compare RCTs by understanding the severity of disease in patients entering studies.

Percentile benchmarks also have an important role in individual patient care, and can answer the question, "How is my patient doing?" Figure 1 represents the percentile benchmarking system applied to a patient with severe RA. It is easy to understand the benefit this patient received from his most recent disease modifying antirheumatic drug.

Thirty five years ago Landsbury described an index that might be considered to be the forerunner of benchmarking systems for rheumatology.²¹ That index was based on experience rather than data, and faded from use as treatments and measures changed. With far better data and valid, sensitive measures, it is now time to turn from relative change scores which are difficult to interpret to easy to understand, meaningful benchmark measurements.

F WOLFE

National Data Bank for Rheumatic Diseases,
Arthritis Research Centre Foundation,
University of Kansas School of Medicine,
Wichita, Kansas, USA

H K CHOI

Arthritis Unit,
Department of Medicine,
Massachusetts General Hospital,
Boston, USA

Correspondence to: Professor F Wolfe, Arthritis Research Centre Foundation, 1035 N Emporia, Suite 230, Wichita, KS 67214, USA
fwolfe@arthritis-research.org

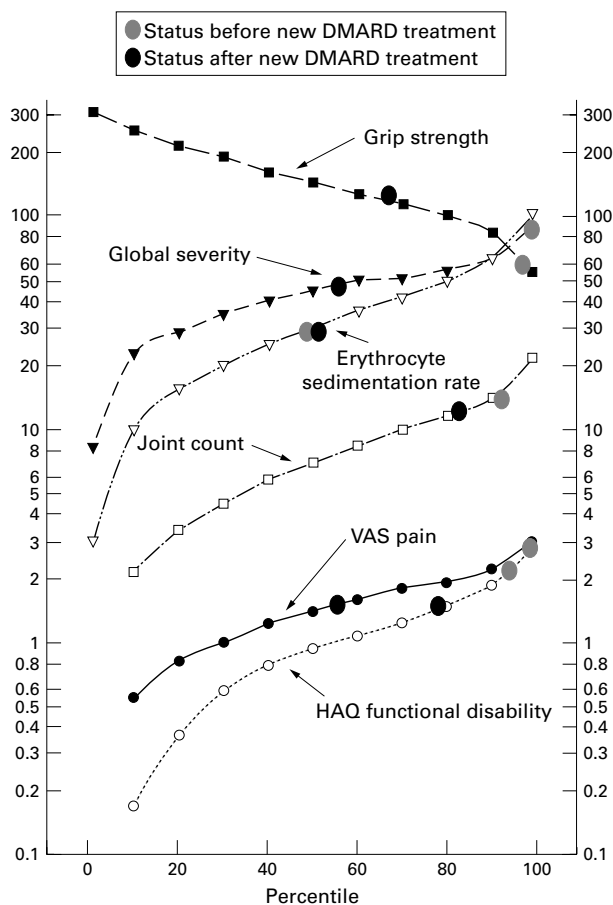


Figure 1 Benchmarking in clinical practice. Patient is a 48 year old attorney who had rheumatoid arthritis (RA) for 18 years. RA was nodular, seropositive, highly erosive, and had required total joint arthroplasty of elbows, hip, and left knee. Previous treatment included methotrexate, sulfasalazine, hydroxychloroquine, penicillamine, azathioprine, etanercept, and prednisone (7.5 mg ongoing). The figure shows the status and change in status after six months' treatment using benchmarking by percentiles.

- 1 Hamill PV, Drizd TA, Johnson CL, Reed RB, Roche AF. NCHS growth curves for children birth-18 years. United States. *Vital Health Stat* 11 1977;i-iv:1-74.
- 2 Wolfe F, Hawley DJ. The development of a disease status index (DSI) for rheumatoid arthritis [abstract]. *Arthritis Rheum* 1992;35(suppl):S219.
- 3 Lassere M, Wells G, Tugwell P, Edmonds J. Percentile curve reference charts of physical function: Rheumatoid arthritis population. *J Rheumatol* 1995;22:1241-6.
- 4 Wolfe F. The prognosis of rheumatoid arthritis: assessment of disease activity and disease severity in the clinic. *Am J Med* 1997;103:12S-18.
- 5 Wolfe F, Hawley DJ, Pincus T. Standard values for self-report questionnaires: assessments in 20,574 rheumatic disease patients [abstract]. *Arthritis Rheumatism* 1999;42(suppl):S75.
- 6 Fries JF, Spitz PW, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
- 7 Ware JE, Kosinski M. *The SF-36® physical and mental health summary scales: a manual for user's of version 1*. 2nd ed, Lincoln: QualityMetric, 2001.
- 8 Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. *Med Care* 1988;26:724-35.
- 9 Wiles NJ, Scott DGI, Barrett EM, Merry P, Arie E, Gaffney K, et al. Benchmarking: the five year outcome of rheumatoid arthritis assessed using a pain score, the Health Assessment Questionnaire and the Short Form-36 (SF-36) in a community and a clinic based sample. *Ann Rheum Dis* 2001; 60:956-61.
- 10 Wiles NJ, Lunt M, Barrett EM, Bukhari M, Silman AJ, Symmons DP, et al. Reduced disability at five years with early treatment of inflammatory polyarthritis: results from a large observational cohort, using propensity models to adjust for disease severity. *Arthritis Rheum* 2001;44:1033-42.
- 11 Harrison B, Symmons D. Early inflammatory polyarthritis: results from the Norfolk Arthritis Register with a review of the literature. II. Outcome at three years. *Rheumatology (Oxford)* 2000;39:939-49.
- 12 Harrison BJ, Symmons DPM, Barrett EM, Silman AJ. The performance of the 1987 ARA classification criteria for rheumatoid arthritis in a population based cohort of patients with early inflammatory polyarthritis. *J Rheumatol* 1998;25:2324-30.
- 13 Wolfe F. A reappraisal of HAQ disability in rheumatoid arthritis. *Arthritis Rheum* 2000;43:2751-61.
- 14 Wolfe F, Pincus T. The level of inflammation in rheumatoid arthritis is determined early and remains stable over the longterm course of the illness. *J Rheumatol* 2001;28:1817-24.
- 15 Barrett EM, Scott DG, Wiles NJ, Symmons DP. The impact of rheumatoid arthritis on employment status in the early years of disease: a UK community-based study. *Rheumatology (Oxford)* 2000;39:1403-9.
- 16 Wolfe F, Hawley DJ. The longterm outcomes of rheumatoid arthritis: work disability: a prospective 18 year study of 823 patients. *J Rheumatol* 1998;25:2108-17.
- 17 Wolfe F, Mitchell DM, Sibley JT, Fries JF, Bloch DA, Williams CA, et al. The mortality of rheumatoid arthritis. *Arthritis Rheum* 1994;37:481-94.
- 18 Wolfe F, Zwillich SH. The long-term outcomes of rheumatoid arthritis: a 23-year prospective, longitudinal study of total joint replacement and its predictors in 1,600 patients with rheumatoid arthritis. *Arthritis Rheum* 1998;41:1072-82.
- 19 Lipsky PE, van der Heijde DMFM, Stclair EW, Furst DE, Breedveld FC, Kalden JR, et al. Infliximab and methotrexate in the treatment of rheumatoid arthritis. *N Engl J Med* 2000;343:1594-602.
- 20 Moreland LW, Schiff MH, Baumgartner SW, Tindall EA, Fleischman RM, Bulpitt KJ, et al. Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial. *Ann Intern Med* 1999;130:478-86.
- 21 Landsbury J. Methods of evaluating rheumatoid arthritis. In: Hollander JL, ed. *Arthritis and allied conditions*. Philadelphia: Lea and Febiger, 1966:269-91.